开放式荟萃分析的规范化报告

刘宇[1]，陈树铨[2]，樊富珉[1]，邸新[3], 范会勇[4], 封春亮[5], 郭双双[1]，甘怡群[6]，李会杰[7,8]，吕小康[9]，任志洪[10]，徐鹏飞[11], 袁博[12], 左西年[13,14*], 胡传鹏[15*]

[1]清华大学心理学系，北京 100084
[2]Department of Clinical and Counseling Psychology, Teachers College, Columbia University, New York City 10025, USA
[3] Department of Biomedical Engineering, New Jersey Institute of Technology, Newark, NJ 07102, USA
[4]渤海大学教育与体育学院, 锦州 121013
[5]华南师范大学心理学院, 广州 510631
[6]北京大学心理与认知科学学院，行为与心理健康北京市重点实验室，北京 100871
[7]中国科学院行为科学重点实验室(中国科学院心理研究所)，北京 100101
[8]中国科学院大学心理学系，北京 100049
[9]南开大学周恩来政府管理学院社会心理学系, 天津 300350
[10] 华中师范大学心理学院, 武汉 430079
[11] 北京师范大学心理学部, 北京 100875
[12] 宁波大学心理学系暨研究所, 宁波 315211
[13] 北京师范大学认知神经科学与学习国家重点实验室/IDG 麦戈文脑科学研究院, 北京 100875
[14] 国家基础学科公共科学数据中心, 北京 100190
[15] 南京师范大学心理学院, 南京 210023

*** 通讯作者：胡传鹏（hu.chuan-peng@nnu.edu.cn），左西年（xinian.zuo@bnu.edu.cn）**

摘要：荟萃分析（又称元分析）以量化方式综合实证研究结果，是积累整合证据并推进基础与应用研究的重要工具。在开放式科学这种新型研究实践下，荟萃分析是整合开放数据、进行科学推断的重要手段。但荟萃分析的过程繁复、方法多样，其执行上的严谨性对荟萃分析质量影响较大，因此报告的透明性和开放性对其结果的可重复性尤为重要。为了解中文荟萃分析报告规范性的现状，让更多研究者了解荟萃分析报告的规范性与透明性，提升荟萃分析报告的质量，本文以国际学界广泛使用的系统综述和荟萃分析优先报告条目

（PRISMA）以及开放式科学理念为基础，结合基于效应量的荟萃分析报告特点，拟定《开放与可重复荟萃分析的报告清单》（Preferred Reporting Items for Open and Reproducible Meta-analysis, PRIOR-MA），并以心理学中的荟萃分析为例，对近五年国内发表的 68 篇中文荟萃分析的方法和结果部分进行回顾。基于结果，我们建议荟萃分析报告需要在以下方面进一步加强其报告的开放性与透明性：文献搜索的时间及限制、研究筛选和数据收集过程的细节、文章筛选的流程图、效应量转换的细节、单个研究偏倚状况的评估等。本文为开放式荟萃分析提供了一个较为全面的报告清单（PRIOR-MA），为未来荟萃分析研究提供参考。

关键词：荟萃分析, 报告规范, 可重复性, 质量评价, 开放式科学**引言**

荟萃分析，英文 meta-analysis，另译“元分析”[1]，是针对某一领域内出现的研究问题，对先前实证研究进行量化综合的研究手段，是研究者对该领域的进展进行量化评估、迅速掌握关于该领域进展的重要信息源。 荟萃分析在循证研究中备受重视。荟萃分析一词最早源于心理学家 Glass 对心理治疗研究有效性之间差异的探讨[1]，但广泛应用于实证研究的各个领域。以“meta-analysis”为关键词在 PubMed 数据库中进行检索，仅选择“meta-analysis”这一文章类型，可发现近年来每年发表的荟萃分析数量超过一万五千篇。在中国知网 CNKI 数据库中的检索也发现，近期每年荟萃分析的数量也有三千余篇，增长迅速（见图 1）。在神经信息学领域，荟萃分析也起着巨大的作用。例如，大规模国际合作项目 ENIGMA（Enhancing Neuro Imaging Genetics Through Meta-Analysis）联盟即是通过对多站点的神经影像数据进行荟萃分析的方式来理解正常人脑的功能与结构，也为理解精神疾病的神经基础提供了大量的实证数据。同样，中国研究者也开始领导这类大规模的荟萃分析合作项目。例如，抑郁症静息态功能磁共振多中心数据荟萃分析（REST-Meta-Major Depressive Disorder，REST-Meta-MDD）项目[2-4]。荟萃分析提供的丰富数据和更有说服力的结论可以为理论和实践带来巨大影响。在开放式科学这种新型实践下，荟萃分析是整合开放数据，从而进行更合理推断的重要手段，越来越多的研究者相信荟萃分析中的证据累积思维是未来研究的重要原则[5]。

1 荟萃分析报告的规范性

作为一种重要的研究方法，荟萃分析的研究思路相对简单清晰，但其质量依赖于研究方法上的严谨性和规范性。从已经发表的研究来看，国内研究者可能对这些规范的关注不够。通过在 CNKI 检索，我们发现自 2009 年 PRISMA 声明发表以来，2010 年~2019 年共发表 24147 篇中文荟萃分析，然而提及这一国际主流荟萃分析报告规范的荟萃分析文章仅有 48 篇[2]，也就是说，大部分国内发表的荟萃分析在实施过程中可能并未参考报告规范。尽管越来越多研究者意识到通过发表荟萃分析论文能对某一领域带来较大影响，然而未按规范实施的荟萃分析却可能带来结论可靠性低，或难以重复的不良后果[6]。研究者需要加强对荟萃分析规范的认识，同时，荟萃分析报告规范迫切需要进行中国化，被更多国内研究者熟知。这种必要性在我国当前需要提升科学产出质量的背景下，显得尤为突出。

1 本文将 meta-analysis 翻译为“荟萃分析”基于如下考虑：英文“Meta-analysis”有将众多研究经筛选聚集后进行分析之意。先前采用“元分析”的译法借用了心理学中“meta-cognition”等术语的翻译，即某一主体对自己的认识和分析，但汉字中“元”字本身并无类似用法。因此“元分析”更像西方科学文化中的外来词。反之，“荟萃”一词在现代汉语词典中有“会集”、“聚集”之意，且“萃”字作为动词用于“萃取”等情境。因此，本文采用更贴近中国文化的“荟萃分析”一词。

2 我们以“PRISMA”为被引文献，并以“元分析”或“荟萃分析”为关键词，在 CNKI 中对中文期刊进行检索后发现检索结果为 51，排除 2 篇综述类与 1 篇英文文献，最后结果为 48 篇，最后检索日期为 2020 年 12 月 25 日。

图 1 不同年份发表的中英文荟萃分析论文数量。在 PubMed 数据库中以“meta-analysis”作为关键词进行搜索，选择“meta-analysis”为文章类型，共检索到 135721 条结果；在 CNKI 数据库中以“元分析”、“meta-analysis”分别为关键词，并排除含有混淆词“有限元分析”、“边界元分析”、“QTL 元分析”和“求积元分析”的文章，共检索到 26536 条中文结果。两个数据库检索覆盖时间为 1976 年~2019 年（为更清晰展示，排除 1990 年以前的 3 篇论文），检索日期：2020 年 11 月 26 日。

Fig. 1 Number of meta-analysis studies published in Chinese and English over the years. For meta-analyses published in English, we searched in the PubMed database, used “Meta-analysis” as the key word, restricted the article type to “Meta-analysis”, and found 135721 items. For meta-analyses published in Chinese, we searched in the CNKI database, used “元分析” (Chinese meaning is meta-analysis) and “meta-analysis” as key words and excluded items that contain “Finite element method”, “Boundary element analysis”, “QTL meta-analysis” and “Quadrature element analysis”, we found 26536 items. Time range of the literature searching is from 1976 to 2019. Date of article search was November 26, 2020.

为了解当前中文荟萃分析报告规范性的现状，以及让更多研究者了解荟萃分析报告的规范性，提升荟萃分析的质量，本文尝试以心理学为例，对荟萃分析报告规范进行中国化，并以此为标准对已发表的中文荟萃分析进行初步评估。具体而言，我们首先以广泛使用的 PRISMA 为基础，结合心理学荟萃分析中主流方法的特点，并参考当前国际心理学界开放科学的趋势[7, 8]，通过深入讨论，形成中文版的荟萃分析报告清单。其次，我们根据中文版的荟萃分析报告清单中方法与结果部分条目，对国内近五年心理学荟萃分析报告现状进行评估，以了解当前中文荟萃分析报告中存在的问题。最后，我们将结合评估结果和当前开放式科学的趋势，有针对性地提供建议，借此希望国内研究者能够更合理地使用荟萃分析方法。

1.1 国际学界关于荟萃分析报告的规范性

荟萃分析作为一种综合已有数据的方法，受关注较多的集中于其数据来源和质量。例如，“同质性问题（mixing apples and oranges）”、“出版偏见（publication bias, 也叫 file drawer problem, 抽屉箱问题）”和“数据质量”的讨论由来已久[9]。但值得注意的是，荟萃分析在文献选择、纳入、编码以及结果解释上存在许多的决策和操作空间，并直接影响到荟萃分析的数据来源和质量，从而影响最终结论。也就是说，荟萃分析实施中的严谨性和透明程度直接关系到其结论的可靠性和可重复性[10]。随着荟萃分析文章数量的激增，可重复性问题也日益突显：针对同一研究主题的荟萃分析，结论却存在差异，有时甚至完全对立[11]。例如 Ferguson 和 Kilburn 就暴力电子游戏与暴力行为之间的关系进行了一项荟萃分析，结论认为两者之间不存在关联[12]；然而，在随后出版的针对同一主题另一项荟萃分析的结论却完全相反，认为攻击性行为的增加与暴力游戏存在强烈的正相关[13]。类似争议让

研究者开始反思荟萃分析的可靠性与可重复性，关注荟萃分析的规范性，以增强其严谨性和透明性。Gøtzsche、Hróbjartsson、Marić 和 Tendal 按照原作者描述的方法对医学领域 27 篇荟萃分析数据转换后的标准化平均差异（standardized mean differences）进行检验与重复，结果发现 10 篇（37%）荟萃分析无法重现该效应量，其中 7 篇存在明显错误，1 篇荟萃分析效应重复后不再显著，另外 2 篇则变为显著[14]。Maassen 等试图重复来自 33 篇心理学荟萃分析的 500 个主要效应量，但基于荟萃分析所报告信息，224 个效应量无法重复，其原因大部分是由于这些荟萃分析文章对数据的选择、效应量计算过程等方法信息报告不完整[6]。

上述例子表明，荟萃分析结论的可靠性与可重复性严重依赖于其报告的规范性。但研究者可能对这些规范性的重视程度不够。事实上，荟萃分析在规范上的严重问题早已引发一些研究者的讨论。在上世纪八十年代，Sacks 等对 86 篇使用英语发表的荟萃分析从 6 个重要方面（包括 23 个条目）进行了详细评估，其结果表明，86 篇荟萃分析中仅有 24 篇（28%）完整地报告了 6 个重要方面的内容[15]。针对同一主题在 1999 年更新的研究发现，荟萃分析报告质量并未随时间推移而有所改善[16]，近期研究也得出类似结论[17]。

为解决荟萃分析报告规范性的问题，一些研究者和学术组织开始尝试制定荟萃分析的报告规范或指南。由于荟萃分析过程繁复、方法多样所导致的过度灵活性，报告规范的主要作用在于按照荟萃分析的大致流程，指出需要报告的方法细节，从而让同行能够检查或者重复这些步骤，增强分析的可重复性。报告规范旨在起到三个重要的作用：第一，帮助研究者提前意识到在荟萃分析过程中所涉及的各项元素，指导其实施荟萃分析；第二，帮助审稿人、期刊编辑迅速地对荟萃分析稿件质量进行检查；第三，帮助后续研究者重复该荟萃分析或者对相似的荟萃分析进行比较。要起到这些作用，一个好的报告规范应该包括三个重要的方面：推荐价值、指导作用与成为荟萃分析文章出版的参考标准。也就是说一个好的报告规范，应能在方法上严谨而规范，促进报告的透明程度，避免研究中的偏倚（bias，也译为偏见、偏差）；具有被研究者推荐的价值；其有详细的说明与指导，能够对研究者实施该研究进行有效的帮助；同时这些信息有助于形成一系列具体的准则与标准成为期刊和学术组织制定相关审稿政策或者标准时的参考基础。

医疗卫生领域专家率先组成国际性小组并于 1999 年制定了首个报告规范 QUOROM（Quality of Reporting of Meta-analyses statement）[18]。2009 年该规范得到进一步修订，更名为 PRISMA 系列报告规范（Preferred Reporting Items for Systematic Reviews and Meta-Analyses）[19]。该规范包括 7 个部分，27 个条目清单和一个四阶段的流程图，与其它报告规范相比，PRISMA 具有两个方面的优势：一方面，PRIMSA 规范得到 Centre for Reviews and Dissemination、Cochrane Collaboration、Council of Science Editors、National Evidence-based Healthcare Collaborating Agency (NECA)与 World Association of Medical

Editors 5 个国际组织与几百种期刊支持；另一方面，PRISMA 能够对研究者提供切实有效的指导，其不仅涵盖了大部分随机对照试验，而且针对每一个条目，均有完备与清晰的阐述与解释性文件，并配有具体的例子[20]。尽管目前难以比较参考报告规范与未参考报告规范的荟萃分析在质量上的差异，但从已有的研究来看，如果作者按照 PRISMA 要求进行报告，会有更多的主观决策过程得以公开，这将增强该荟萃分析的重复性。已有研究已经发现，在 PRISMA 发表后，有更多的文章提到 PRISMA 中的相关概念与方法（例如，更多使用 PRISMA 推荐的流程图），文章总体质量呈上升趋势[21]；如果荟萃分析中提到了 PRISMA，它们报告的信息往往也会更加全面[22]。当荟萃分析中报告了更加全面的信息时，那么也更可能被重复出来[6]。

此外，从引用量上来看，PRISMA 是目前引用最多的报告规范：自 2009 年发表以来，截止到 2020 年 10 月 13 日搜索，PRISMA 的两篇介绍性文章[19, 20]在 Google 学术上的引用量分别为 26198 和 66667 次。因此，PRISMA 系列规范形成了相对完备的规范体系，是比较成熟的荟萃分析规范。美国心理学会（American Psychological Association, APA）也做过类似的尝试[23, 24]，但其影响与被引用度低于 PRISMA（自 2008 年推出以来，MARS 与更新版本在 Google 学术上仅被引用 170 次与 150 次，检索日期：2020 年 10 月 13 日）。

1.2 对荟萃分析报告规范的中国化尝试

如前所述，国内研究者对荟萃分析报告规范关注和使用较少，对报告质量的评估也鲜有报道。为了让研究者能够直接参考中文荟萃分析报告规范，本团队对目前国际上的荟萃分析报告规范进行中国化，整合出适合心理学同时对其他荟萃分析具有参考意义的报告清单。

根据上文所述，整合荟萃分析报告清单时主要遵循两个原则：第一，符合国际主流的荟萃分析报告规范，即该清单能够满足荟萃分析报告规范的三个重要方面（推荐价值、指导作用和出版要求）；第二，能够体现当前研究中的新动向，即开放与透明的研究实践需求[7, 25]。遵循第一个原则，本文介绍的报告规范以 PRISMA 为基础。如上所述，相较于其他报告规范，PRISMA 达到了好的报告规范的三个方面要求，同时具有很高的关注度与认可度。但由于 PRISMA 主要应用于医学研究，直接使用 PRISMA 对于心理学或其他领域荟萃分析研究者来说并不满足“指导手册”的要求。因此，本文在 PRIMSA 的基础之上，整合最新版 MARS 中的相关条目。遵循第二个原则，本文参考目前被众多杂志所采用的《提升透明与开放指南》（Transparency and Openness Promotion, TOP, https://cos.io/our-services/top-guidelines/）[7]，加入对透明与开放的要求。虽然 PRISMA 最近开始要求对荟萃分析方案进行预注册[26]，但其并未要求对荟萃分析中涉及到的数据与分析代码进行开放。因此，本荟萃分析报告规范将 TOP 规范中相关的要求采纳进来。

通过本文专家组的讨论与整合，最终形成《开放与可重复荟萃分析的报告清单》（Preferred Reporting Items for Open and Reproducible Meta-analysis, PRIOR-MA，核心方法部分见表 1，完整版以及核心方法部分解释见存放于科学数据存储库（ScienceDB）中的补充材料[27]）。PRIOR-MA 力图为常规的、基于效应量的荟萃分析提供一个报告清单，以期为国内研究人员撰写和报告荟萃分析提供借鉴与参考。本清单共有 30 个条目，其中方法部分的 14 个条目（条目 5 至条目 18）是其核心内容。在这 14 个条目中，12 条为基础要求，均以 PRISMA 为基础修订，2 条为开放式科学的新要求（表 1 中带*的条目 17 与条目 18，均为 PRIOR-MA 的新增条目），即鼓励研究者报告。

表 1 PRIOR-MA（方法部分）

Table 1. PRIOR-MA (method section)

章节/主题	#	清单条目
方法		
方案预注册	5	陈述是否提前对荟萃分析研究方案进行预注册。如果预注册过，则需要提供相关信息（例如，公开预注册的网址）。
文献纳入标准	6	陈述文献纳入的标准： 6.1 变量：自变量（例如，实验的操作，处理、干预或预测变量的类型），因变量（例如，对临床研究中存在的潜在积极与消极效应进行综合后的结果）或相关研究中涉及的变量； 6.2 研究设计的标准（例如，样本抽样的方法或实验处理的方式）； 6.3 针对同一研究或同一样本存在多次报告或使用同一样本进行多次测量的情况，不仅需要陈述纳入标准的研究特征（例如，自变量、因变量和研究设计等），同时需要陈述报告特征（检索时间、语种、发表日期和发表状态等），以确定哪个研究需要进行报告； 6.4 研究纳入中存在的限制（例如，研究的年代、语言、地区或进行报告的方式上存在可能因此产生偏倚的困难）； 6.5 文献纳入标准的变化。预注册时提交的文献纳入标准是否在研究开始后发生改变，或者研究过程中文献纳入标准是否发生改变，例如增加或减少纳入研究的变量，作者需要报告发生了什么改变，以及这些改变是什么时间发生，原因是什么；

		6.6 如何处理文献中缺失信息，包括与纳入标准的信息相关信息的缺乏（例如，缺乏关于研究设计的信息）及与数据分析相关信息的缺失（例如，没有报告关于结果的量化数据）。
信息来源	7	描述搜索得到所有信息来源（例如，包括数据库中获取的文献信息，联系作者后获取的额外信息）和最后一次搜索日期。
搜索过程	8	至少完整呈现一个主要使用的数据库在线数据搜索策略，包括限定的范围，搜索时使用的关键词、搜索代码等，以确保搜索的可重复性。
研究筛选	9	陈述每一步研究筛选过程，包括： 9.1 每一步筛选文献的决策方式（如，通过阅读题目、摘要或全文）； 9.2 筛选过程中执行者的胜任力与人数。如果是一位执行者，如何评估提取数据过程的信度问题；如果有多位执行者，如何处理多人进行筛选可能存在的不一致问题。
数据收集过程	10	描述从报告中提取数据的方法，包括： 10.1 作为摘录数据依据的变量及其种类； 10.2 数据提取过程中执行者的胜任力与人数。如果是一位执行者，如何评估提取数据过程的信度问题；如果有多位执行者，如何处理多人提取可能存在的不一致问题。
数据条目	11	列出并定义纳入文献中所有与荟萃分析相关信息（例如，文献的作者、发表年份、被试信息、研究设计或控制组信息等）。
单个研究偏倚风险	12	描述如何评估纳入的研究本身可能存在的偏倚（说明这种评估是在研究方法过程层面还是结果层面，还是两者都有），以及这种偏倚如何在结果综合中进行处理。
描述性统计指标	13	描述使用的效应量指标（比如，相关系数、Cohen’s *d*），包括用来计算单个研究效应量公式。如果同一研究存在多个效应量可能引发的非独立性问题（例如，同一研究中存在多个实验处理、多个控制组、多个相关系数或多个指标量等产生的非独立效应量），陈述处理的方法。
结果综合	14	描述进行跨研究效应量综合和对不同研究的效应进行异质性评估的方法（如固定效应模型或随机效应模型的稳健方差估计（robust variance estimation）），具体包括：

		14.1 选取某一综合方法的原因； 14.2 对研究结果进行加权的方法； 14.3 估计单个研究与总体结果变异的方法（如，置信区间）； 14.4 描述所有的转换/校正（例如，小样本或非等组样本数据）、调整（例如，聚合，数据缺失，或建构不同水平间的关系）以及对这些进行的数据验证）； 14.5 如果进行了贝叶斯分析，报告先验分布的选择与模型拟合的评估。 注：（如采用更加复杂或者新颖的方法，可以进行适当地调整）
跨研究偏倚风险	15	评估所有可能导致跨研究证据累积中产生偏倚的可能性（例如，发表偏倚，选择性报告）。
额外分析	16	如果进行了额外分析，需要描述使用的方法（例如，亚群体分析，敏感性分析，元回归分析），指出哪些是预先计划的。
数据可获得性*	17	荟萃分析中使用的数据（即从文献中提取的数据条目）是否公开。如将荟萃分析的原始数据存放在一个公开并可长期存放的在线数据库中，方便其他研究者进行重复与在此研究基础上的累积。
分析过程可获得性*	18	荟萃分析中使用的软件名称及版本号。荟萃分析的代码是否公开，如果公开，在哪可以获得（例如：与数据存在同一在线数据库），方便其他研究者对分析过程的重复。

*为 PRISMA 2009 中未提及的开放科学新要求，鼓励研究者进行报告。

从表 1 中可以看到，在方法部分，PRIOR-MA 考虑了荟萃分析中可能存在过度灵活性的方面：纳入荟萃分析的文献（条目 6、7、8、9）、作为荟萃分析“原料”的数据（条目 10、11、12、13）、进行荟萃分析的统计模型（条目 14、15、16）。其中条目 12 和 15 是偏倚评估，而条目 5、条目 17 和条目 18 则主要关注透明性和数据开放，以保证其可重复性。关于方法部分条目的更多内容与阐释请见线上补充材料[27]。

2. 国内心理学荟萃分析文献报告现状

我们以心理学为例，对已发表的荟萃分析文章进行检索和评估，以了解国内荟萃分析报告的现状以及需要加强之处。本文以 PRIOR-MA 为基础，选取荟萃分析核心的方法与结

果部分，对近五年中文心理学期刊上发表的传统荟萃分析进行评估。评估的目的在于以心理学中文期刊为例，了解当前中文荟萃分析报告的情况，而非评价期刊或文章的质量。由于荟萃分析的重要性，不少研究者均曾对其领域的荟萃分析质量进行过评估[17, 28-39]。为避免误解，本文对所有纳入的荟萃分析文章进行匿名化处理。

2.1 研究方法

文献纳入标准　根据研究目的，本文纳入标准为：（1）荟萃分析类研究，纳入传统以效应量综合为目的的荟萃分析，排除非效应量综合的荟萃分析，以避免方法与结果报告中部分条目不适用带来的混淆；（2）使用中文发表在主流心理学期刊上，排除非心理学期刊上的荟萃分析文献；（3）选取完整已出版的文献，排除增刊、未公开发行与会议摘要文献；（4）发表于近五年的文献（2015~2020）。本纳入标准存在的如下限制：语言仅包括中文，文献来源仅包括期刊，但这些限制与本报告的目的有关。

信息来源　所有文献来自中国知网数据库（cnki.net）和《心理学报》、《心理科学》、《心理科学进展》、《心理发展与教育》、《心理学探新》、《心理与行为研究》、《中国临床心理学杂志》七本期刊官方网站数据，文献覆盖时间为 2015 年 1 月 1 日至 2020 年 1 月 1 日，初次检索时间为 2018 年 4 月 12 日，最后一次检索时间为 2020 年 1 月 15 日。

搜索过程　以“系统综述”、“Meta 分析”、“荟萃分析”、“元分析”、“Meta-analysis” 作为关键词进行检索，通过如下代码在 CNKI 上进行搜索：

文献来源=('心理学报'或'心理科学'或'心理科学进展'或'心理发展与教育'或'心理学探新'或'心理与行为研究'或'中国临床心理学杂志') 和题名=('系统评价'或'系统综述'或'Meta 分析'或'元分析'或'Meta-analysis'或'荟萃分析')

共检索到 88 文献。整个检索过程由作者一独立完成。作者七使用相同关键词在同样的数据库与期刊上进行独立的检索，最终与作者一进行核对。文献搜索过程符合本研究的目的，无纳入标准之外的限制。

研究筛选　通过文献题目与摘要，由作者一与作者七独立根据文献合格规范进行筛选。遇到分歧时与通讯作者进行讨论并处理分歧。

数据收集过程与编码　我们首先根据 PRIOR-MA 内容，进一步细化条目，编制出编码手册供评分参考（见补充材料 2）。其次是建立数据收集表格，表格中需要纳入文献中研究的基本特征，如题目、作者、发表时间与期刊来源等，以及需要进行编码并评估的内容。数据收集由作者一与作者七独立完成，并对收集内容一致性进行讨论与评估，无法统一时与通讯作者进行进一步讨论并做出决策。

编码指南中的每一具体评分条目需要被编码为“1”，“0”，“0.5”或“？”。“1”表示该条目完全报告，编码为“0”表示该条目完全未报告，“0.5”表示该条目未完全报告，“？”表示该条目在编码过程中很难判断。编码为不清楚的条目需要由作者一与作者七进行讨论并做决策，如无法决定时，继续与通讯作者进行讨论并最终决定。为使决策过程更加透明，我们在有争议的条目得分中增加了描述来澄清原因。

编码过程由作者一按照编码指南对所有纳入的荟萃分析文献进行独立编码，作者七随机从每年发表的荟萃分析中抽取4~6篇荟萃分析文献进行独立编码（共26篇），两者编码完成后计算评价者一致性Kappa值[40]（计算过程见补充材料3[27]）。Mchugh认为Kappa作为一致性指标，相较于百分比，更好地将偶然存在的一致性情况加以考虑，避免对一致性进行高估[40]。根据先前的文献，我们对于Kappa值的判断标准为：0.40~0.59之间为一致性好，在0.60~0.74之间为相当好，0.75及以上为一致性非常好[41]。

数据条目 文献信息摘取条目包括：（1）文献题目；（2）文献发表时间；（3）文献发表期刊；（4）文献作者；（5）文中涉及PRIOR-MA方法与结果部分的条目（条目5-条目25）。

描述性统计指标 按照PRIOR-MA，本文对纳入文献研究方法与结果部分，共32个具体条目内容进行评估。通过报告状况百分比（即，每个条目“完整报告”、“部分报告”和“完全未报告”三种情况文章数量分别占文章总数的百分比）来评估单个条目内容在所有纳入荟萃分析中完成的情况。

结果综合 本文仅进行系统综述描述性统计，无荟萃分析过程，未进行结果综合。

单个研究偏倚与跨研究偏倚 本文主要目的是按照PRIOR-MA，了解国内心理学荟萃分析的完成状况，不以单个研究自身的研究目的为评估内容，主要存在的偏倚来自发表偏倚。

数据可获得性与分析过程可获得性 由于对纳入评估文献原作者的尊重，仅将评分结果上传进行保存并公开，见：补充材料4[27]。

2.2 结果

2.2.1 研究筛选与研究特征

初步检索共获得文献88篇，经两位作者进行独立搜索与筛选，有6篇文献不符合纳入标准（4篇综述，2篇非中文写作），阅读全文后进一步筛选再次排除14篇文献（10篇文献主要采用时间序列方法，4篇为神经成像荟萃分析方法），最终共有68篇荟萃分析纳入评估。筛选流程及结果见图2。

纳入荟萃分析均为中文，并发表于国内本领域杂志中，根据研究目的，样本特征未公开。感兴趣的读者可以与作者联系，了解纳入荟萃分析的具体信息。

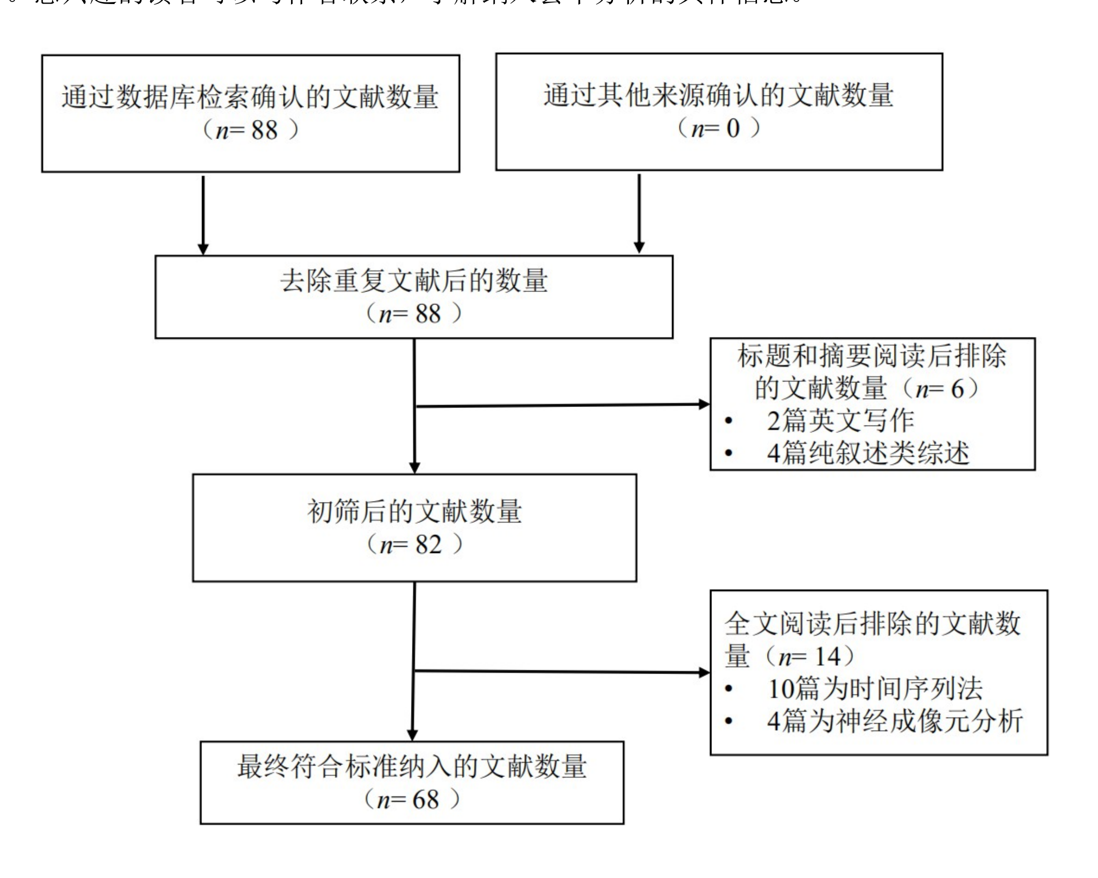


图 2 文献筛选流程图

Fig. 2 Flow diagram of article screening

2.2.2 评分一致性

对作者一与作者七独立编码的分数计算 Kappa 值，结果为 0.814，表明两名评分者的一致性达到了较高的水平。

2.2.3 总体报告质量

基于 PRIOR-MA 的荟萃分析质量评估，两位编码者针对编码不同的条目进行再次讨论，最终统一编码得分。本文报告的结果为最终统一编码得分。

21 个条目和 32 个具体评分内容报告情况见图 3。条目 5 中“预注册情况”，条目 6 中“报告特征限制”，条目 8 中“搜索过程中的限制”，以及“数据与分析过程可获得性”两个条目平均分为 0 分，即所有荟萃分析研究均未提及相关内容。其他容易被忽视的内容还包括：“最后搜索时间”，“筛选人数情况”，“研究质量评估”，以及“数据转换方

法”等内容，完整报告百分比均未超过 30%。

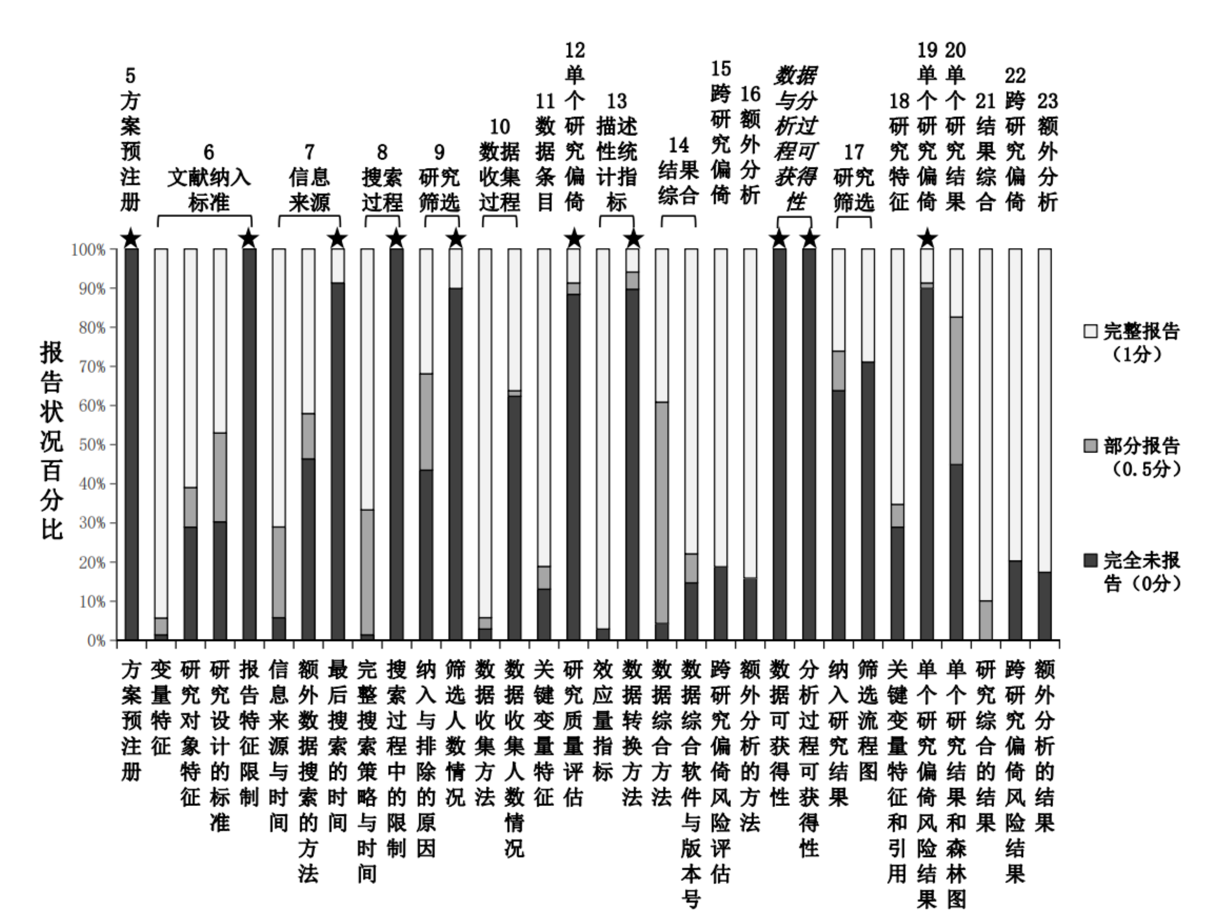


Fig. 3 Methods and results reporting of meta-analyses in Chinese psychological journals, as evaluated by PRIOR-MA. Items 5-23, on the top of the bar plot, are items from PRISMA and PRIOR-MA, the italics are new items from PRIOR-MA; items below the bar plot are further specified items in the coding manual. Bars with ★ are items that were severely under-reported (< 30% articles).

完全报告超过 90%的条目仅 4 项，包括条目 13“描述性统计指标”中的效应量指标、条目 14 和条目 21“结果综合”部分。表明国内研究在大部分的条目上都较为缺乏关注。在“综合测量指标”与“结果综合”部分完成情况较好。但需要注意的是，条目 14 中“数据整合方法”部分内容完整报告文章数量为 39.1%，超半数的文章未在方法部分对荟萃分析使用的模型进行描述，或未明确告知所选择的方法。总体而言，国内荟萃分析完整报告的条目比例较低，具有较大的提升空间。

3. 讨论

荟萃分析是重要的实证方法，完整和透明地报告荟萃分析过程，有助于提升荟萃分析

的可靠性和可重复性。关注荟萃分析报告的严谨性和可重复性，对于实证研究中的积累具有重大意义。本文尝试以 PRISMA 为基础，整合形成中文版荟萃分析报告清单 PRIOR-MA，为国内同行提供参考。同时，本文参考 PRIOR-MA 的核心条目（方法与结果部分的条目），以国内心理学近五年荟萃分析报告为例进行现状调查，以期有针对性地提供建议调查结果表明，使用 PRIOR-MA 对荟萃分析报告情况进行评估时，不同的评分者之间具有较好的评估者一致性，表明 PRIOR-MA 能够较好地被使用。通过对已发表的中文荟萃分析进行调查，我们发现 PRIOR-MA 对荟萃分析的可靠性和可重复性以及报告质量上，在以下几个方面可能存在显著提升：

第一， **文献纳入过程**

荟萃分析的第一步在于全面查找已有的研究并纳入与荟萃分析范围相符的文献。从搜索文献到最终纳入的文献之间，存在着多个步骤，并且均存在一定的主观性和自由度。对 68 篇中文荟萃分析的调查表明，尽管所有荟萃分析文章均有提及文献纳入标准、信息来源、搜索过程与研究筛选，但仍然有一些细节易被忽视。

在文献纳入标准中最容易忽视的内容是“研究报告特征限制”，68 篇荟萃分析均未提及相关内容。PRISMA 的文献纳入标准条目指出，荟萃分析需要报告纳入研究的“研究特征（study characteristics）”与“报告特征(report characteristic)”(Liberati et al.,2009)。前者在心理学中应包括纳入研究的被试群体，自变量与因变量等特征；而后者涉及研究使用的语言、文献发表年代和发表状态（例如，未公开出版或正在出版）等，这些因素可能在制定纳入标准时存在许多限制。例如，有研究表明，使用不同语言的研究纳入荟萃分析后会产生截然不同的结果[26]，即使作者纳入的研究语言仅为英语或中文，也有必要报告是否存在纳入其他语言研究的可能性，而非默认为某一种语言。荟萃分析研究者对研究特征可能的限制因素需要保持敏感性，尽量消除对后续研究者带来的误解。

在明确文献纳入标准后，作者需要根据计划收集相关信息。这一阶段涉及对“信息来源”与“搜索过程”的报告。我们发现很少有文献对“最后一次数据库文献搜索日期”进行报告（91.3%未报告）。报告搜索每个数据库开始与结束的时间是非常重要的，因为通过不同的时间标签搜索到不同的文献，会直接影响到荟萃分析最终的结果[42]。

在搜索过程报告部分，纳入调查的荟萃分析文章大部分对搜索使用的关键词与数据库进行报告，但极少提及搜索执行中可能存在的语言等限制，这一点与先前的研究结果一致[43]。然而，几乎所有的搜索均存在局限性，例如在时间、资金支持、数据库购买、因语言或搜索技巧而无法获取的信息[20]。报告搜索中的限制一方面让读者能够衡量当前荟萃分析结果的可靠性，另一方面也能让作者对可能存在的偏倚提高警惕。

在研究筛选上主要有两点需要关注：一是关于筛选过程的清晰性。PRISMA 推荐使用筛选流程图来描述每一步纳入与排除研究的过程[19, 20]，其原因是它能用简洁的图形和数字让读者对筛选过程一目了然，快速获取大量信息[44]。然而，本调查中的荟萃分析文章，大部分仅呈现最终纳入研究的数量，仅少数研究使用筛选流程图（28.9%完整报告）。这可能导致读者难以快速清晰地获知研究是如何进行筛选的信息；二是在筛选中许多研究未提及参与筛选的人数，筛选者胜任力以及如何处理筛选一致性，或者如何避免一位筛选者可能存在的错误等问题（仅有 10.1%的研究进行完整报告）。一位研究者进行筛选有可能由于过于主观或者粗心而出现错误，有研究表明，文献筛选过程也是一个极其困难的决策过程，一位研究者很难独立胜任[44]。

第二，　**荟萃分析中数据质量的控制**

确定了纳入的文献之后，荟萃分析的研究者需要从文献中提取数据并进行分析。这些步骤在 PRIOR-MA 中有两点需要注意：一是数据收集与编码过程的清晰报告，二是对纳入荟萃分析的单个实证研究质量进行评估。

在数据收集上，数据收集过程的清晰报告可以有效减少错误与偏倚[45]。但数据的收集本身可能会面临着诸多的选择（例如，对同一研究假设，可能对应着多个统计结果，研究者可能有多种做法）。这种情况下，是否有多名研究者来独立进行数据收集，以及数据收集的一致性问题就非常重要。我们的调查发现，大约四成的荟萃分析文章报告了数据收集人数与如何解决数据收集不一致的问题，表明部分研究者已经意识到这个问题的重要性。但仍然有超过半数据的研究者未报告这方面信息。另外，在基于多站点的荟萃分析中，数据收集往往由发起者向合作者征集，而非从文献中摘取。但在数据比较复杂时，对同一假设，不同站点可能会采用不同的分析流程和不同的方法去分析数据[46]。因此，保证不同站点之间数据分析的方式与流程的一致性变得十分重要，例如，REST-Meta-MDD 中统一使用 dpabi 软件进行数据分析[47]。

荟萃分析存在的偏倚主要分为单个研究偏倚与跨研究偏倚。我们的调查发现，大部分荟萃分析对于跨研究的偏倚有较高完成度（81.2%完整报告）。但是很少有研究提及单个研究的偏倚（仅 8.7%完整报告）。也就是说，作者未报告纳入荟萃分析的研究本身质量如何，衡量这些研究中可能存在的偏倚，并针对这些偏倚采取相应的措施（如调整权重等）。虽然大部分荟萃分析文章是对已经发表的文献进行分析，可以假定这些荟萃分析中的原研究已经过了同行评议的审核，无重大的偏差。但是，仍然需要注意的是，由于心理学近年来的可重复危机[25, 48-51]，在纳入荟萃分析时研究者可能需要回答：（1）每个研究在多大程度上是可靠的？（2）该研究在哪些方面存在偏倚？（3）发现的偏倚是否会影响总体结论[17]？一个比较全面的评估需要从研究的内部效度、外部效度、构念效度与统计结论效度四

个方面对原研究进行简要的质量评估[52]。此外，一些文献质量评估方法也有助于降低纳入研究质量带来的偏倚，例如，GRADE 证据评价系统[53]和 Jadad 量表[54]。同样，在医学领域荟萃分析中，尤其是随机分组试验，对单个研究报告的质量会从多个维度进行评估，目前已有许多工具来评估不同类型的研究[55]。

第三，　**荟萃分析的数据分析与结果部分**

获得数据原材料后，荟萃分析需要通过对不同研究中摘取出来的数据进行转换，便于对主要研究问题的效应量进行合并。由于目前实证研究的发表中并未采用统一的格式进行数据报告，因此当研究者进行荟萃分析时，需要通过对文献中的统计数据对效应量进行转化或者估计[6]。当荟萃分析中文章数量较大时，这一过程非常繁杂，涉及公式选取，合理性的探讨和软件使用等步骤。在调查的 68 篇荟萃分析中，仅有 4 篇文章报告转换所需的公式与过程，占总数的 5.9%。如上所述，在荟萃分析实施的过程中，研究者选择的效应量转换公式过于灵活性，且涉及主观决策部分。因此，如果未在荟萃分析文章中对转换过程进行详细报告，后续研究者很难对荟萃分析研究的结果进行重复验证[6]。此外，荟萃分析中的数据收集与效应量转换过程的规范性也提示我们，在未来的开放式科学平台的建设中，需要整合结果报告的标准和规范。当研究者提供可以机读 (machine readable) 结果的数据，将大大简化未来荟萃分析的过程，且降低荟萃分析过程中的灵活性和主观性。

在效应量整合过程中，尽管只有 4.3%的文章未提及整合的模型与方法，但是许多荟萃分析文章的报告并不完整（部分报告占总数的 56.5%）。主要存在两方面问题，一是模型的选定较为简单，大部分研究直接选用随机效应模型，而非根据研究本身数据特征进行分析，对模型选择的原因进行有逻辑的推理说明；另一方面容易忽视的是整合所用的工具，包括软件与版本号。目前荟萃分析过程在市面上存在许多可供选择的软件，这些软件由于种类或自身版本差异，可能导致不同的操作逻辑与结果差异。

在结果部分需要注意的是对单个研究结果的报告，呈现单个研究结果有助于读者检查每个研究的特征，评估不同研究之间结果变异，同时有助于数据摘取错误发生后的回查。在所评估的文章中仅有 17.4%进行完整报告，其中大部分研究未使用 PRISMA 推荐的森林图（forest plot）[20]。森林图的使用能将纳入研究中的效应量、置信区间以及百分比权重等重要信息清晰呈现给读者，有助于读者快速了解荟萃分析的全貌[56]。

第四，荟萃分析的透明与开放性

研究方案预注册、数据与分析过程可获得性是提升研究质量的实践，近年来已经成为国际上实证研究的新趋势，尤其是在神经信息领域对其开放科学的研究实践取得了显著进展[8]。PRISMA 也将会在新一版本中加入开放式科学的元素（见其预印本[21]）。在医学卫

生领域的研究中，对荟萃分析的方案进行注册已经是国际惯例，其中最著名的是 Cochrane 荟萃分析数据库。Cochrane 协作组早在 1993 年成立之初，就要求作者对荟萃分析进行注册，并提交计划方案。后续研究也证实其必要性。Silagy 等发现，在 Cochrane 荟萃分析数据库中 47 篇已发表的论文，其中 43 篇与计划书相比存在较大改动[57]。Kirkham 等通过对连续 3 期 Cochrane 中的系统评价进行评估，结果发现改变原有的计划方案，会对研究结论造成偏倚和错误的解释[58]。如果未进行预注册，后续研究者则难以获知原作者在选择性报告上是否存在问题。但由于 Cochrane 荟萃分析制作周期相对较长，允许范围限于临床试验相关的荟萃分析注册，且注册后需要在 Cochrane 中发表，因此预注册这一前瞻性的做法并未得到普及。为了填补 Cochrane 协作组在注册方面的空缺，另一个使用较为广泛的是 YORK 大学建立的 PROSPERO（International Prospective Register of Ongoing Systematic Reviews.www.crd.york.ac.uk/PROSPERO）系统综述国际预注册平台，它允许临床试验以及与医疗健康相关的动物实验荟萃分析注册，且未规定注册的荟萃分析在何处发表。PROSPERO 为研究者们发表方案和信息共享提供了极大的便利[59]。随着方案预注册的便利性的提升，研究方案的透明性、准确性与完整性也变得极为重要，也出现了专门为预注册而定制报告条目的规范 PRISMA-P（Preferred Reporting Items for Systematic Reviews and Meta-Analyses protocols）[26]。

除了方案预注册外，荟萃分析数据与代码的公开也有助于增强荟萃分析的可重复性。在神经影像荟萃分析中，REST-Mata-MDD 项目已经将荟萃分析使用的统计结果数据公开（见[2-4, 8]），也有专门用来存放各类数据的 NeuroVault[60]和特定荟萃分析结果数据存储的 Archive of Neuroimaging Meta-analyses (ANIMA)[61]。随着对荟萃分析可重复性的重视[62]，一些杂志要求按照《提升透明与开放指南》（Transparency and Openness Promotion, TOP, https://cos.io/our-services/top-guidelines）[7]，加入对透明与开放的要求。

除上述内容外，我们还发现作者在写作时可能会将结果部分内容放入研究方法部分，最常出现的是将研究筛选结果、研究特征等内容放入研究方法部分。这样的报告方式可能让读者无法清晰地获取相应的信息，容易造成遗漏与误解。

以上的分析表示，通过参考荟萃分析的报告规范，能够有效地提高荟萃分析的质量。值得提出的是，当研究方法不断深入与扩展，报告规范往往会滞后于当前研究的需求，需要长期不断的更新与完善。例如，神经成像领域发展出了激活似然估计法（Activation Likelihood Estimation, ALE）、核密度分析法（Kernel Density Analysis, KDA）与多水平核密度分析法（Multilevel Kernel Density Analysis, MKDA）[63]，不同于传统基于效应量分析的荟萃分析方法[64, 65]。因此，为方便国内心理学研究者进行自检或者审稿的清单，并结合当前国际上报告规范的最新发展，本文的在线补充材料中提供了完整版的《开放与可重复荟萃分析的报告清单》（Preferred Reporting Items for Open and Reproducible Meta-analysis,

PRIOR-MA)，该清单共 30 个条目，包括一篇荟萃分析报告的完整内容，能够为研究者在撰写和报告荟萃分析时提供参考。当然，由于荟萃分析研究方法的复杂性与多样性，我们也鼓励研究者查阅 PRISMA 声明的解释性文件[20]及其他资料以得到更详细的条目信息。

4. 总结与展望

在涉及到繁复数据分析流程的研究领域，形成报告规范有助于整个领域的健康发展，如人类脑图谱学会（The Organization for Human Brain Mapping）专门就 fMRI/MRI 的数据分析与分享推荐了最佳实践指南[66]。同样，在荟萃分析领域，一个旨在增强开放式和可重复的报告规范，可能有助于提升荟萃分析的质量，进而促进科学知识的有效积累。基于此目的，本文尝试通过对国际广泛使用的 PRISMA 进行中国化，对条目进行详细说明并采用近期发表的开放式荟萃分析文章作为样例进行分析，以保证其可操作性[27]。一方面，希望未来更多的期刊编辑、审稿人和读者能够了解荟萃分析报告的规范，以 PRIOR-MA 或类似规范作为参考，促进审稿和荟萃分析的质量评估；另一方面，也希望 PRIOR-MA 或者类似的规范能够对同行有启发意义，在未来出现的更多具体荟萃分析方法中，发展出更加适合特定类型的荟萃分析报告规范或者检查清查。

同时需要指出的是，本文所提出的报告清单有其局限性。第一，PRIOR-MA 仅代表本文作者团队根据其知识和经验进行的整合，难免存在不足之处，未来可能需要进一步修改和改进。第二，PRIOR-MA 主要针对基于效应量所进行的传统荟萃分析，尤其在方法部分，未完全覆盖近期出现的一些荟萃分析新方法，如网络荟萃分析或者神经成像中的 ALE 方法等。未来研究可以基于类似原则，提出有助于提升报告质量与可重复性的更具体荟萃分析报告规范。

最后，本文所提出的荟萃分析报告清单，主要针对传统以人工编码为基础的荟萃分析。当前科学出版中数字化技术日益成为主流，本文的调查也可以为未来开放式共享数据平台的建设提供启示：未来通过联合政府机构、学术组织/机构、期刊出版方和研究者多方的力量，对科学研究的结果报告中相对基础和常规的内容进行标准化。例如，对于有公开数据的文章，文章中的结果部分与数据平台上的数据进行挂钩；对于无法公开数据的文章，可以邀请作者提供可以机读的数据结果。这种标准化的结果报告，将方便后续研究者进行检索和效应量的转换，大大提高荟萃分析的效率和准确性，减少人工实施荟萃分析出错的可能性。在不久的将来，这类自动化荟萃分析方式将极大可能加快科学研究中去伪存真和知识积累的速度。

参考文献

[1] Glass G V. Primary, secondary, and meta–analysis of research [J]. Educational researcher, 1976, 5: 3–8
[2] Schmaal L, Pozzi E, Ho T C, et al. ENIGMA MDD: seven years of global neuroimaging studies of major depression through worldwide data sharing [J]. Translational Psychiatry, 2020, 10: 1–19
[3] Xia M, Si T, Sun X, et al. Reproducibility of functional brain alterations in major depressive disorder: Evidence from a multisite resting–state functional MRI study with 1,434 individuals [J]. Neuroimage, 2019, 189: 700–714
[4] Yan C–G, Chen X, Li L, et al. Reduced default mode network functional connectivity in patients with recurrent major depressive disorder [J]. Proceedings of the National Academy of Sciences, 2019, 116: 9078–9083
[5] Simons D J, Holcombe A O, Spellman B A. An introduction to registered replication reports at perspectives on psychological science [J]. Perspectives on Psychological Science, 2014, 9: 552–555
[6] Maassen E, van Assen M A, Nuijten M B, et al. Reproducibility of individual effect sizes in meta–analyses in psychology [J]. PloS one, 2020, 15: e0233107
[7] Nosek B A, Alter G, Banks G C, et al. Promoting an open research culture: author guidelines for journals could help to promote transparency, openness, and reproducibility [J]. Science, 2015, 348: 1422–1425
[8] Zuo X N, Zang Y F, Gao J H. Toward neuroinformatics of neuroimaging data sharing and open brain science. In Journal of SCIENTIA SINICA Vitae, 2020, https://www.sciengine.com/doi/10.1360/SSV-2020-0292 [左西年, 臧玉峰, 高家红. 推进神经影像数据共享与开放式脑科学 [J]. 中国科学：生命科学, 2020, https://www.sciengine.com/doi/10.1360/SSV-2020-0292]
[9] Sharpe D. Of apples and oranges, file drawers and garbage: Why validity issues in meta–analysis will not go away [J]. Clinical psychology review, 1997, 17: 881–901
[10] Ferguson C J. Comment: Why meta–analyses rarely resolve ideological debates [J]. Emotion Review, 2014, 6: 251–252
[11] de Vrieze J. The metawars [M]. American Association for the Advancement of Science. 2018, 361: 1184–1188
[12] Ferguson C J, Kilburn J. The public health risks of media violence: A meta–analytic review [J]. The Journal of pediatrics, 2009, 154: 759–763
[13] Anderson C A, Shibuya A, Ihori N, et al. Violent video game effects on aggression, empathy, and prosocial behavior in Eastern and Western countries: A meta–analytic review [J]. Psychological bulletin, 2010, 136: 151–173
[14] Gøtzsche P C, Hróbjartsson A, Marić K, et al. Data extraction errors in meta–analyses that use standardized mean differences [J]. Journal of the American Medical Association, 2007, 298: 430–437
[15] Sacks H S, Berrier J, Reitman D, et al. Meta–analyses of randomized controlled trials [J]. New England Journal of Medicine, 1987, 316: 450–455
[16] McAlister F A, Clark H D, van Walraven C, et al. The medical review article revisited: has the science improved? [J]. Annals of internal medicine, 1999, 131: 947–951
[17] Gates N J, March E G. A neuropsychologist's guide to undertaking a systematic review for publication: Making the most of PRISMA guidelines [J]. Neuropsychology Review, 2016, 26: 109–120
[18] Moher D, Cook D J, Eastwood S, et al. Improving the Quality of Reports of Meta–Analyses of Randomised Controlled Trials: The QUOROM Statement [J]. Oncology Research and Treatment, 2000, 23: 597–602
[19] Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta–analyses: the PRISMA statement [J]. PLoS med, 2009, 6: e1000097
[20] Liberati A, Altman D G, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta–analyses of studies that evaluate health care interventions: explanation and elaboration [J]. Journal of clinical epidemiology, 2009, 62: e1–e34

[21] Page M J, McKenzie J, Bossuyt P, et al. Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement [J]. Journal of Clinical Epidemiology, 2020, 134: 103–112
[22] Leclercq V, Beaudart C, Ajamieh S, et al. Meta–Analyses Indexed In Psycinfo Had A Better Completeness Of Reporting When They Mention Prisma [J]. Journal of Clinical Epidemiology, 2019, 115: 46–54
[23] A. P. A. Publications Communications Board Working Group On Journal Article Reporting S. Reporting standards for research in psychology: Why do we need them? What might they be? [J]. The American Psychologist, 2008, 63: 839–851
[24] Appelbaum M, Cooper H, Kline R B, et al. Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report [J]. American Psychologist, 2018, 73: 3–25
[25] Zhu Ying."Open science, data sharing, software sharing", are you ready?. Advances in Psychological Science, 2016, 24: 995–996 [朱滢. "开放科学 数据共享 软件共享", 你准备好了吗? [J]. 心理科学进展, 2016, 24: 995–996]
[26] Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta–analysis protocols (PRISMA–P) 2015 statement [J]. Systematic reviews, 2015, https://doi.org/10.1186/2046-4053-4-1
[27] Liu Y, Chen S Q, Fan F M, et al. Supplementary Materials for A checklist for reporting meta–analysis in open science era [M]. Science Data Bank. 2021, http://www.dx.doi.org/10.11922/sciencedb.00339
[28] Peters J P, Hooft L, Grolman W, et al. Reporting quality of systematic reviews and meta–analyses of otorhinolaryngologic articles based on the PRISMA statement [J]. PLoS One, 2015, 10: e0136540
[29] Schalken N, Rietbergen C. The reporting quality of systematic reviews and meta–analyses in industrial and organizational psychology: a systematic review [J]. Frontiers in psychology, 2017, 8: 1395
[30] Tan W K, Wigley J, Shantikumar S. The reporting quality of systematic reviews and meta–analyses in vascular surgery needs improvement: a systematic review [J]. International Journal of Surgery, 2014, 12: 1262–1265
[31] Wen J, Ren Y, Wang L, et al. The reporting quality of meta–analyses improves: a random sampling study [J]. Journal of clinical epidemiology, 2008, 61: 770–775
[32] Willis B H, Quigley M. The assessment of the quality of reporting of meta–analyses in diagnostic research: a systematic review [J]. BMC medical research methodology, 2011, https://doi.org/10.1186/1471-2288-11-163
[33] Yao L, Sun R, Chen Y–L, et al. The quality of evidence in Chinese meta–analyses needs to be improved [J]. Journal of clinical epidemiology, 2016, 74: 73–79
[34] Zhi X, Zhang Z, Cui J, et al. Quality of meta–analyses in major leading orthopedics journals: a systematic review [J]. Orthopaedics & Traumatology: Surgery & Research, 2017, 103: 1141–1146
[35] LI B, Zhang D Q, Wangg X Q, et al. Quality Evaluation of China's Sports Intervention Meta-analysis Report Based on PRISMA Statement. Sport Science Research, 2020, 41: 85–89 [李博, 张丹青, 王雪强, 等. 基于 PRISMA 声明的我国运动干预类元分析报告质量评价 [J]. 体育科研, 2020, 41: 85–89]
[36] Li J, Yao Y S, Jin Y L, et al. A systematic evaluation on the quality of Meta-analysis in articles published in the Chinese Journal of Epidemiology. Chinese Journal of Epidemiology, 2013, 34: 819–825 [李杰, 姚应水, 金岳龙, 等.《中华流行病学杂志》55 篇 Meta 分析文献的质量评价 [J]. 中华流行病学杂志, 2013, 34: 819–825]
[37] Su L, Li Y Q, Kou L H, et al. Standardized report quality analysis of Meta-analysis articles and systematic reviews published in Chinese nursing journals in 2018. Chinese Nursing Research, 2019, 33: 1866–1871 [苏琳，李亚琴，寇丽红, 等. 2018 年我国护理期刊刊载 Meta 分析与系统评价文献规范化报告质量分析. 护理研究, 2019, 33(11), 1866–1871]
[38] Zhang W X, Xiong Y, Xu C, et al. Methodological and reporting quality of dose-response meta-analysis published in Chinese journals. Chinese Journal of Evidence-Based Cardiovascular Medicine, 2017: 9: 1300–1304 [张维欣, 熊莺, 徐畅, 等. 中文期刊发表的剂量 - 反应 Meta 分析方法学质量及报告质量现状 [J]. 中国循证心血管医学杂志, 2017, 9: 1300–1304]

[39] Zhou P X, Yan Y Y, Zhai S D. Reporting and methodological quality assessment for systematic reviews/meta-analyses conducted by hospital pharmacists in China. Chinese Journal of Evidence-Based Cardiovascular Medicine, 2017, 17: 228–234 [周鹏翔，闫盈盈，翟所迪. 国内医院药学人员系统评价/Meta 分析的方法学与报告质量评价. 中国循证心血管医学杂志, 2017, 17: 228–234]
[40] McHugh M L. Interrater reliability: the kappa statistic [J]. Biochemia medica: Biochemia medica, 2012, 22: 276–82
[41] Orwin R G, Vevea J L. Evaluating coding decisions [J]. The handbook of research synthesis and meta–analysis, 2009, 2: 177–203
[42] Shojania K G, Sampson M, Ansari M T, et al. How quickly do systematic reviews go out of date? A survival analysis [J]. Annals of internal medicine, 2007, 147: 224–33
[43] Egger M, Zellweger–Zähner T, Schneider M, et al. Language bias in randomised controlled trials published in English and German [J]. The Lancet, 1997, 350: 326–9
[44] Vu–Ngoc H, Elawady S S, Mehyar G M, et al. Quality of flow diagram in systematic review and/or meta–analysis [J]. PloS one, 2018, 13: e0195955
[45] Jones A P, Remmington T, Williamson P R, et al. High prevalence but low impact of data extraction and reporting errors were found in Cochrane systematic reviews [J]. Journal of clinical epidemiology, 2005, 58: 741–742
[46] Botvinik–Nezer R, Holzmeister F, Camerer C F, et al. Variability in the analysis of a single neuroimaging dataset by many teams [J]. Nature, 2020, 1–7
[47] Yan C G, Wang X D, Zuo X N, et al. DPABI: Data Processing & Analysis for (Resting–State) Brain Imaging [J]. Neuroinformatics, 2016, 14: 339–351
[48] Baker M. 1,500 scientists lift the lid on reproducibility [J]. Nature, 2016, 7604: 452–454
[49] Schooler J W. Metascience could rescue the 'replication crisis' [J]. Nature, 2014, 515: 9
[50] Hu C P, Wang F, Guo J C S, et al. The replication crisis in psychological research. Advances in Psychological Science, 2016, 24: 1504–1518 [胡传鹏, 王非, 宋梦迪, 等. 心理学研究中的可重复性问题: 从危机到契机 [J]. 心理科学进展, 2016, 24: 1504–1518]
[51] Jiao C, Zhang M Q. The Lost Boundary:A Study of the Null Hypothesis Testing Method in Psychology. Social Sciences in China, 2014, 2: 148–163 [焦璨, 张敏强. 迷失的边界: 心理学虚无假设检验方法探究 [J]. 中国社会科学, 2014, 2: 148–163]
[52] Valentine J C. Judging the quality of primary research [J]. The handbook of research synthesis and meta–analysis, 2009, 2: 129–146
[53] Alonso–Coello P, Schünemann H J, Moberg J, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction [J]. The British Medical Journal, 2016, 353: i2016
[54] Brouwers M C, Johnston M E, Charette M L, et al. Evaluating the role of quality assessment of primary studies in systematic reviews of cancer practice guidelines [J]. BMC medical research methodology, 2005, 5: 8
[55] Zeng X, Zhang Y, Kwong J S, et al. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review [J]. Journal of evidence–based medicine, 2015, 8: 2–10
[56] Lewis S, Clarke M. Forest plots: trying to see the wood and the trees [J]. The British Medical Journal, 2001, 322: 1479–80
[57] Silagy C A, Middleton P, Hopewell S. Publishing protocols of systematic reviews: comparing what was done to what was planned [J]. Journal of the American Medical Association, 2002, 287: 2831–2834
[58] Kirkham J J, Dwan K M, Altman D G, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews [J]. The British Medical Journal, 2010, 340: c365
[59] Booth A, Clarke M, Dooley G, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews [J]. Systematic reviews, 2012, 1: 1–9
[60] Gorgolewski K J, Varoquaux G, Rivera G, et al. NeuroVault. org: A repository for sharing unthresholded statistical maps, parcellations, and atlases of the human brain [J]. Neuroimage, 2016, 124: 1242–1244
[61] Reid A T, Bzdok D, Genon S, et al. ANIMA: A data–sharing initiative for neuroimaging meta–analyses [J]. Neuroimage, 2016, 124: 1245–1253

[62] Braver S L, Thoemmes F J, Rosenthal R. Continuously cumulating meta–analysis and replicability [J]. Perspectives on Psychological Science, 2014, 9: 333–42
[63] Wager T D, Lindquist M, Kaplan L. Meta–analysis of functional neuroimaging data: current and future directions [J]. Social cognitive and affective neuroscience, 2007, 2: 150–158
[64] Eickhoff S B, Laird A R, Fox P M, et al. Implementation errors in the GingerALE Software: description and recommendations [J]. Human brain mapping, 2017, 38: 7–11
[65] Hu C P; Di X, Li J W, et al. Meta-analysis of Neuroimaging Studies. Advances in Psychological Science, 2015, 23: 1118–1129 [胡传鹏, 邸新, 李佳蔚, 等. 神经成像数据的元分析 [J]. 心理科学进展, 2015, 23: 1118–1129]
[66] Nichols T E, Das S, Eickhoff S B, et al. Best practices in data analysis and sharing in neuroimaging using MRI [J]. Nature neuroscience, 2017, 20: 299–303

A standardized checklist on reporting meta-analysis in the open science era

LIU Yu[1], CHEN Shuquan[2], FAN Fumin[1], DI Xin[3], FAN Huiyong[4], FENG Chunliang[5], GUO Shuangshuang[1], GAN Yiqun[6], LI Huijie[7,8], LV Xiaokang[9], REN Zhihong[10], Xu Pengfei[11], YUAN Bo[12], Zuo XiNian[13,14*] & HU Chuan-Peng[15*]

[1] *Department of Psychology, Tsinghua University, Beijing* 100084, *China;*
[2] *Department of Clinical and Counseling Psychology, Teachers College, Columbia University, New York City* 10025, *USA;*
[3] *Department of Biomedical Engineering, New Jersey Institute of Technology, Newark* 07102, *USA;*
[4] *College of Education Sport Science, Bohai University, Jinzhou* 121013, *China;*
[5] *School of Psychology, South China Normal University, Guangzhou* 510631, *China;*
[6] *School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, School of Psychological and Cognitive Sciences, Peking University, Beijing* 100871, *China;*
[7] *Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing* 100101, *China;*
[8] *Department of Psychology, University of Chinese Academy of Sciences, Beijing* 100049, *China;*
[9] *Department of Social Psychology, Zhou Enlai School of Government, Nankai University, Tianjin* 300071, *China;*
[10] *School of Psychology, Central China Normal University, Wuhan* 430079, *China;*
[11] *Faculty of Psychology, Beijing Normal University, Beijing 100875, China;*
[12] *Department of Psychology, Ningbo University, Ningbo* 315211, *China;*
[13] *State Key Laboratory of Cognitive Neuroscience and Learning/IDG McGovern Institute for Brain Research, Beijing Normal University, Beijing* 100875, *China;*
[14] *National Basic Science Data Center, Beijing* 100190, *China;*
[15] *School of Psychology, Nanjing Normal University, Nanjing* 210023, *China*

Abstract: Meta-analysis is a crucial tool for accumulating evidence in basic and applied research. In the open science era, meta-analysis becomes an important way for integrating open data from different sources. Meanwhile, because of researchers' high degree of freedom introduced by multiple-steps and multiple-choices in each step of meta-analysis, the openness and transparency are crucial for reproducing the results of meta-analysis. To reflect the transparency and openness of meta-analysis reports published in Chinese journals and improve the transparency and openness of future meta-analysis by Chinese researchers, we developed a Chinese version of a checklist for meta-analysis (Preferred Reporting Items for Open and Reproducible Meta-analysis, PRIOR-MA), which was based on the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA) and the principle of openness and transparency. We then surveyed the methods and results of 68 meta-analysis papers published in mainstream Chinese psychological journals in the last five years. Our results revealed that the openness and transparency of Chinese meta-analysis reports need to be improved, especially in the following aspects: the date/time and limitation of literature search, the details of screening and data collection, the flow chart of article screening, the details of effect size transformation, and the evaluation of individual research bias. The checklist, which lists almost all aspects that an open meta-analysis should include, can be used as a guide for future meta-analysis. This paper provides

a comprehensive checklist (PRIOR-MA) for open meta-analysis, and provides a reference for future meta-analytic research.

Keywords: meta-analysis, reporting standard, reproducibility, quality of reporting, open science